TwoPath U-Net for Automatic Brain Tumor Segmentation from Multimodal MRI data

Keerati Kaewrak[1] , John Soraghan[1], Gaetano Di Caterina[1] and Derek Grose[2]

[1] Centre for Signal and Image Processing, Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, United Kingdom
{keerati.kaewrak, j.soraghan, gaetano.di-caterina}@strath.ac.uk
[2] Beatson West of Scotland Cancer Centre, Glasgow, United Kingdom
Derek.Grose@ggc.scot.nhs.uk

Abstract. A novel encoder-decoder deep learning network called TwoPath U-Net for multi-class automatic brain tumor segmentation task is presented. The network uses cascaded local and global feature extraction paths in the down-sampling path of the network which allows the network to learn different aspects of both the low-level feature and high-level features. The proposed network architecture using a full image and patches input technique was used on the BraTS2020 training dataset. We tested the network performance using the BraTS2019 validation dataset and obtained the mean dice score of 0.76, 0.64, and 0.58 and the Hausdorff distance 95% of 25.05, 32.83, and 37.57 for the whole tumor, tumor core and enhancing tumor regions.

Keywords: Brain Tumor, Deep Learning, Segmentation.

1 Introduction

Glioma is a type of brain tumor that abnormally grows from glial cells of the brain and it is the most common type of brain tumor that is found in both adults and children. Early and accurate diagnosis are important keys to patients' survival rate [1]. Because the tumor appearances vary from patient to patient, brain tumor segmentation represents a particular challenging image segmentation problem. MRI is currently the imaging modality of choice for brain tumor assessment because of its superior soft-tissue contrast. The brain tumor segmentation (BraTS) challenge has been publishing the BraTS datasets that are widely used for the brain tumor segmentation study for almost a decade [2]. The dataset provided multimodal MRI scans which were acquired from different MRI sequence setting. The multimodal dataset consists of a T1-weighted (T1) scan, a post-contrast T1-weighted (T1+Gd) scan, T2-weighted (T2) scan and a T2 Fluid Attenuated Inversion Recovery (FLAIR) scan. Since the tumor appears differently on each MRI modality, it is important to use multimodal MRI for brain tumor segmentation [3].

Deep learning based models have been reported and proven to be the *state-of-the-art* method for brain tumor segmentation task in the previous BraTS challenges [2, 4].

Convolutional neural networks using small kernels to extract the essential features from the multimodal MRI data and the technique of increasing the depth of the networks were presented in [5]. Two-pathways CNNs [6] included a cascaded local and global feature extraction paths using different size of the kernels. These local and global feature extractions introduced the idea of giving the network different aspects of the input data to learn. However, these conventional convolutional neural network use fully connected layers as a tumor classification and required post-processing to construct the prediction probability maps of tumor segmentation. U-Net [7] was originally introduced for biomedical image segmentation. The encoder-decoder network architecture that proposed skip connection or concatenation of feature maps from contraction or down-sampling path to the corresponding up-sampling result features in the same block level to provided more precise localization information back into the dense feature maps levels. The up-sampling path of the U-Net restores the dense feature maps back to the original input data dimension with the prediction of the segmentation mask and background of the images. This up-sampling process eliminates the fully connected layers in the conventional convolutional neural network. Furthermore, U-Net is able to give a precise segmentation results using only a few hundreds of annotated training data. Hence, U-Net has become the powerful baseline method for medical imaging segmentation problems including automatic brain tumor segmentation [8–10]. The data augmentation is often used to increase the number of training data [7–9, 11]. However, the previous U-Net based models [9, 10] produced binary segmentation of tumor mask and background of the images. To obtain all tumor sub-regions, the process involves the network training for each sub-regions segmentation separately and then post-processing was applied to obtain all tumor structures segmentation.

In this paper, we present a novel encoder-decoder network architecture that is based on U-Net for multi-class brain tumor segmentation. We replaced the contraction or down-sampling paths of the U-Net with the local and global feature extraction paths inspired by cascaded Two-pathways CNNs and applied the random flipping along axis for data augmentation. We also implemented the proposed network using different input strategies by comparing the segmentation results of full-size images input training and patches input training approach.

The remainder of the paper is organized as follows. Data pre-processing and methodology are presented in Section 2. Experimental results are reported in Section 3. Finally, the discussion is presented in Section 4.

2 Methods

2.1 Data pre-processing and Augmentation

BraTS2020 training dataset consists of multimodal MRI scans from 369 patients that come from 293 high-grade glioma (HGG) and 76 low-grade glioma (LGG) patients. Each patient has 4 MRI scans; T1-weighted (T1), T1-weighted with gadolinium enhancing contrast (T1+Gd), T2-weighted (T2) and FLAIR image volumes. Each type of

the MRI scan obtained by using different MRI sequence setting acquisition. These images were resampled and interpolate into 1x1x1 m^3 with the size of 240x240x155. The data set provides segmentation ground truth annotated by expert neuro-radiologists [2, 12–14]. Annotation label comprises of label 1 for the necrotic and non-enhancing tumor (NCR/NET), label 2 for the peritumoral edema (ED), and label 4 for the enhancing tumor (ET). Fig. 1 illustrates original FLAIR images and overlaid of annotated ground truth labels. Red, yellow and blue contours are NCR/NET, ED and ET tumor regions, respectively.

Data normalization was performed on each MRI scan by subtracting the mean of each MRI scan and dividing by its standard deviation. We also applied random flipping along the axis to increase training data as shown in Fig. 1. We split data into the ratio of 0.9:0.1 for cross-validation during the training phase.

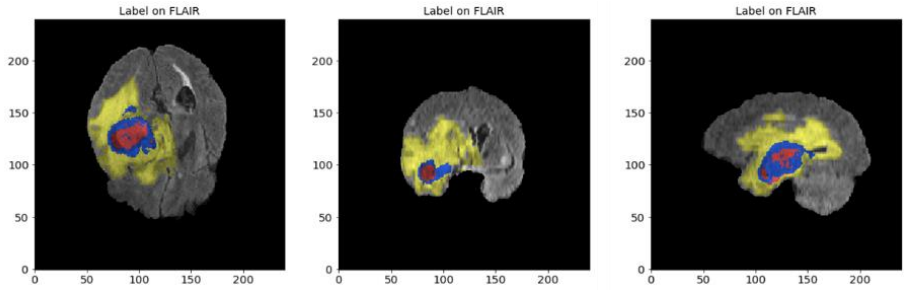

Fig. 1. An example of original image (left) and random flipping results (middle and right) from training dataset with overlaid ground truth labels; NCR/NET (red), ED (yellow), and ET (blue).

2.2 Network Architecture

Fig. 2. shows an illustration of the proposed TwoPath U-Net, which is the encoder-decoder model based on U-Net [7]. It comprises 3 down-sampling blocks, a bottleneck layer and 3 up-sampling blocks with the multimodal MRI input data. Two feature extraction paths which are 3x3 convolution layers with ReLU and 12x12 follows by 9x9 convolution layers with ReLU are used as local and global feature extractions paths. These local and global feature extraction paths capture low-level essential features from the input. The concatenated local and global features provide multi-perspective of the essential context from the input to the model. Then, 2x2 max-pooling operator is performed to halve the size of the feature maps dimension and these features maps become the input of the next down-sampling block. The feature extractions are repeated until the feature maps increase from 4 to 512 dense features. Then two repeated 3x3 convolution layers followed by ReLU are performed in the bottleneck block of the network architecture.

For up-sampling blocks, a 3x3 convolution with bilinear interpolation with stride of 2 is used to double images resolution in both dimensions and to halve the feature maps, followed by 3x3 convolution with ReLU. The result of the up-sampling is then concatenated with the corresponding features from the down-sampling side of the same block

level as shown in Fig. 2. These concatenations or skip connections provide the higher resolution features with local and global localization contextual information to the up-sampling process. The process repeated until the feature resolution increase back to original resolution with 64 feature maps. Finally, 1x1convolution layer [15] of length 1x1x64 with softmax function is employed to produce the output of segmentation mask with 4 classes prediction of all tumor regions.

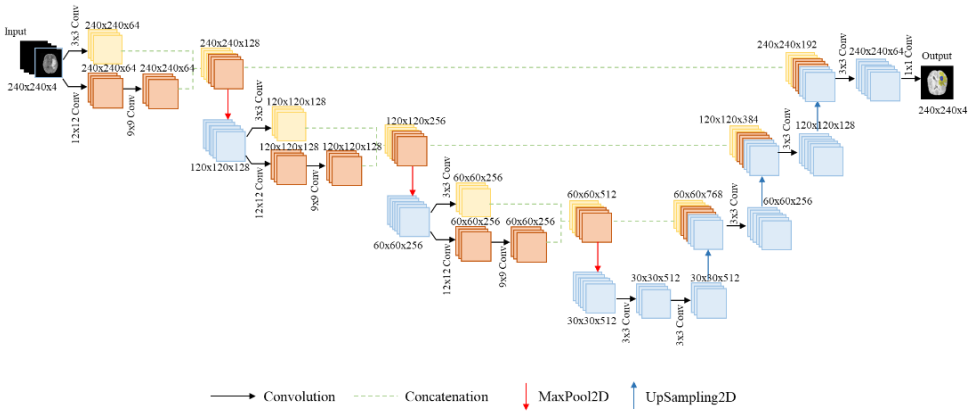

Fig. 2. TwoPath U-Net network architecture.

2.3 Experiments

We performed two experiments using a full-image training and patches training approach. For full-size image training, input image of the size 240x240x4 were fed through 3 blocks of the proposed networks as shown in Fig. 2. At the end of the down-sampling path, the resolution of the input data were halved in both dimension from 240x240 to 30x30 and the feature maps were increase from 4 to 512. For up-sampling path, the dense feature maps resolution from bottleneck of the network were doubled in each block back to original image resolution and final output was the prediction of 4 classes tumor regions segmentation.

Since the majority of the segmentation results are the background of the images. We wanted to eliminate the background of the images that do not contain the brain and the tumor extents during the network training process. Hence, we performed another experiment using patches training. We cropped the original image in to the size of 160x160x4 and then trained the network using cropped input data in the same manner as the full-size images training. The summary of the operations and its output dimension of the patches training approach are illustrated in Table. 1.

Table 1. Model summary for patches segmentation training approach.

Layer	Operations	Output Shape
Input	-	(160,160,4)
Down Sampling 1	12x12 Conv2D+ReLU, 9x9 Conv2D+ReLU, 3x3 Conv2D+ReLU — Concatenation	(80,80,128)
Pool 1	Max Pooling 2D	(80,80,128)
Down Sampling 2	12x12 Conv2D+ReLU, 9x9 Conv2D+ReLU, 3x3 Conv2D+ReLU — Concatenation	(80,80,256)
Pool 2	Max Pooling 2D	(40,40,256)
Down Sampling 3	12x12 Conv2D+ReLU, 9x9 Conv2D+ReLU, 3x3 Conv2D+ReLU — Concatenation	(40,40,512)
Pool 3	Max Pooling 2D	(20,20,512)
Bottleneck	Dropout, 3x3 Conv2D+ReLU, 3x3 Conv2D+ReLU, Dropout	(20,20,512)
Up Sampling 1	UpSampling stride of 2, 3x3 Conv2D+ReLU	(40,40,256)
Skip Connection 1	Up Sampling 1, Down Sampling 3 — Concatenation	(40,40,768)
Up Sampling 2	UpSampling stride of 2, 3x3 Conv2D+ReLU	(80,80,128)
Skip Connection 2	Up Sampling 2, Down Sampling 2 — Concatenation	(160,160,384)
Up Sampling 3	UpSampling stride of 2, 3x3 Conv2D+ReLU	(160,160,64)
Skip Connection 3	Up Sampling 3, Down Sampling 1 — Concatenation	(160,160,192)
Prediction	1x1 Convulution+Softmax	(160,160,4)

2.4 Network Training

An Adam optimizer [16] was used with a learning rate of 0.00001, $\beta_1 = 0.9, \beta_2 = 0.99$, all weights were initialized using a normal distribution with mean of 0 and standard deviation of 0.01, and al biases were initialized as 0. We used training batch size of 16 and trained the network for 50 epochs for each segmentation experiment. Categorical cross-entropy for multiclass segmentation was used as a loss function and can be defined as

$$Loss = -\sum_{n=1}^{N} y_c \log p_c \tag{1}$$

where N is the number of classes which are 4 including the background for this study. y_c is the ground truth of the nth class and p_c is the prediction from softmax function . We implemented the proposed network using Tensorflow and Keras library

on PC equipped with NVIDIA GeForce GTX 1070 GPU, an Intel® Core™ i5-8400 CPU 2.80 GHz processor, 16 GB of RAM.

3 Results

3.1 Evaluation metrics

We reported the Dice Similarity Coefficient which gives the similarity between predicted tumor regions segmentation and ground truth by comparing the overlapped area and can be defined as

$$DSC = \frac{2TP}{FP + 2TP + FN} \tag{2}$$

where TP, FP and FN denote the number of true positive, false positive and false negative counts, respectively.
We also reported the model performance using Hausdorff Distance measure. Given two finite point sets $A = \{a_1, ..., a_p\}$ and $B = \{b_1, ..., b_q\}$, the Hausdorff Distance can be defined as

$$H(A, B) = \max(h(A, B), h(B, A)) \tag{3}$$

$$h(A, B) = \max_{a \in A} \left(\min_{b \in B} (d(a, b)) \right) \tag{4}$$

$$h(B, A) = \max_{b \in B} \left(\min_{a \in A} (d(b, a)) \right) \tag{5}$$

where *d(a,b)* is the Euclidean distance between *a* and *b*. *h(A,B)* is called the directed Hausdorff distance from *A* to *B*, which identifies the point $a \in A$ that is farthest from any point of *B* and measures the distance from *a* to its nearest neighbor in *B*. This means that *h(A, B)* first looks for the nearest point in *B* for every point in *A*, and then the largest of these values are taken as the distance, which is the most mismatched point of *A*. Hausdorff distance *H(A, B)* is the maximum of *h(A, B)* and *h(B, A)*. Hence, it is able to measure the degree of mismatch between two sets from the distance of the point of *A* that is farthest from any point of *B*, and vice versa [17].

3.2 Preliminary results

We used BraTS2019 validation dataset which provided multimodal MRI scans of 125 patients as unknown testing data for our proposed network. The BraTS2019 validation dataset does not provide ground truth. Hence, we uploaded the prediction masks to CBICA image processing portal for BraTS 2019 validation segmentation task and obtained the evaluation results. We evaluated the proposed network performance on 3 segmentation tasks; segmentation of whole tumor region (WT) which is the union of

all annotated tumor labels (ED+NCR/NET+ET), segmentation of the gross tumor core (TC) which is the union of annotated label 1 and 4 (NCR/NET + ET), and segmentation of annotated label 4 (ET) [4].

Table 2 shows that the proposed network with full-images training obtained the segmentation accuracy of mean dice score of 0.57, 0.75 and 0.64, and Hausdorff distance (95^{th} percentile) of 37.56, 25.05 and 32.83 for enhancing tumor, whole tumor and tumor core segmentation tasks. Table 3 shows that the proposed network with middle cropped images training obtained the segmentation accuracy of mean dice score of 0.55, 0.73 and 0.62, and Hausdorff distance (95^{th} percentile) of 39.82, 64.43 and 49.75 for enhancing tumor, whole tumor and tumor core segmentation tasks. Table 4 shows that the proposed network with overlapping cropped images training obtained the segmentation accuracy of mean dice score of 0.50, 0.68 and 0.56, and Hausdorff distance (95^{th} percentile) of 53.10, 66.26 and 59.42 for enhancing tumor, whole tumor and tumor core segmentation tasks. We can also see from Table 2-4 that for 75% quantile, our proposed network has the potential of reaching >80% dice score accuracy for whole tumor and tumor core and >75% dice score accuracy for enhancing tumor regions for all experimental approaches. These results means that with further improvement training strategy, the proposed method could give better performance across whole dataset.

Finally, we compared our experimental results from the proposed network to the segmentation results using the original U-Net in Table 5. We can see that the segmentation results using TwoPath U-Net with full-images training approach achieved the highest dice score for all tumor segmentation tasks. TwoPath U-Net with full-images training approach also obtained lowest Hausdorff distance (95^{th} percentile) which means it obtained the lowest degree of mismatch between the ground truth and the prediction results. An example of the comparison results between the proposed method and ground truth segmentation of all tumor regions from the training dataset is shown in Fig. 3.

Table 2. Segmentation results from full-image training.

Metrics/ Tumor regions	Dice Score			Hausdorff Distance 95%		
	ET	WT	TC	ET	WT	TC
Mean	0.57758	0.7586	0.6414	37.56539	25.04765	32.82901
Median	0.73927	0.80849	0.7432	3.52914	43.00697	9.89949
25% quantile	0.30451	0.70338	0.48014	64.24719	70.00714	69.31666
75% quantile	0.82914	0.87162	0.86007	37.56539	25.04765	32.82901

Table 3. Segmentation results from patches training.

Metrics/ Tumor regions	Dice Score			Hausdorff Distance 95%		
	ET	WT	TC	ET	WT	TC
Mean	0.55216	0.73045	0.61785	39.82449	64.42918	49.75158
Median	0.69698	0.77228	0.71616	25.11967	61.22908	52.33546
25% quantile	0.32106	0.67741	0.44093	4.50848	52.02259	18.62794
75% quantile	0.80972	0.84034	0.84906	72.59666	79.0231	76.84693

Table 4. Segmentation results from overlapping patches training.

Metrics/ Tumor regions	Dice Score			Hausdorff Distance 95%		
	ET	WT	TC	ET	WT	TC
Mean	0.50895	0.67751	0.56053	53.09798	66.26139	59.41933
Median	0.61275	0.73703	0.64553	55.17246	63.97656	60.3034
25% quantile	0.2444	0.58653	0.37928	21.54167	52.20153	40.2368
75% quantile	0.75477	0.80855	0.80325	79.04524	77.13365	79.45817

Table 5. The comparison of segmentation results using different training methods on BraTS 2019 validation dataset.

Metrics/ Tumor regions	Dice Score			Hausdorff Distance 95%		
	ET	WT	TC	ET	WT	TC
Original U-Net	0.37982	0.66885	0.46319	58.26046	64.16464	60.10336
TwoPath U-Net Full-images	**0.57758**	**0.7586**	**0.6414**	**37.56539**	**25.04765**	**32.82901**
TwoPath U-Net middle crop	0.55216	0.73045	0.61785	39.82449	64.42918	49.75158
TwoPath U-Net overlapping crop	0.50895	0.67751	0.56053	53.09798	66.26139	59.41933

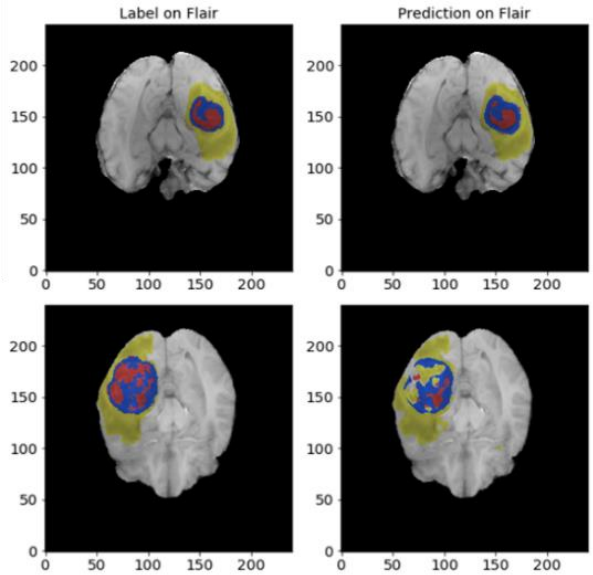

Fig. 3. Comparison between ground truth (left) and prediction (right) segmentation results on original FLAIR images from training dataset with overlaid labels NCR/NET (red), ED (yellow), and ET (blue) regions.

3.3 BraTS 2020 Challenge

We submitted the BraTS 2020 segmentation task using TwoPath U-Net on BraTS 2020 testing dataset. The dataset contains multimodal MRI volumes of 166 patients. The segmentation results validated by the challenge are shown in table 6. We obtained mean dice score of 0.72, 0.66, and 0.64 for the whole tumor, tumor core, and enhancing tumor segmentations. The results from BraTS 2020 Challenge has a similar profile to the preliminary results. At 50% and 75% quantile of dataset, the proposed method achieved over 70% and 80% mean dice score accuracy for all tumor regions segmentation. With an improvement of hyperparameters tuning, the proposed method could perform better across the entire dataset.

Table 6. Segmentation results from BraTS 2020 testing dataset.

Metrics/	Dice Score			Hausdorff Distance 95%		
Tumor regions	ET	WT	TC	ET	WT	TC
Mean	0.63615	0.71819	0.65697	57.982	56.15457	61.4418
Median	0.74136	0.79081	0.78923	18.40394	57.05636	49.08382
25% quantile	0.60252	0.64168	0.55418	4.12311	38.90559	15.141
75% quantile	0.81023	0.85834	0.85119	73.6825	74.62648	76.6368

4 Conclusion

In conclusion, we developed a novel network architecture for fully automatic multi-class brain tumor regions segmentation. The proposed network consists of cascaded local and global feature extraction paths to improve segmentation accuracy. We implemented the proposed network architecture using different input data strategies on the BraTS2020 training dataset and tested the network performance using the BraTS2019 validation dataset as unknown testing data. The proposed network gives better segmentation accuracy than the original U-Net and obtained the mean dice score of 0.76, 0.64 and 0.58 on the validation data and 0.72, 0.66, and 0.64 on the testing data for the whole tumor, tumor core and enhancing tumor regions.

References

1. Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., Ellison, D.W.: The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol. 131, 803–820 (2016). https://doi.org/10.1007/s00401-016-1545-1
2. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, M.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv Prepr. arXiv1811.02629. (2018)
3. Corso, J.J., Sharon, E., Dube, S., El-Saden, S., Sinha, U., Yuille, A.: Efficient Multilevel Brain Tumor Segmentation With Integrated Bayesian Model Classification. IEEE Trans. Med. Imaging. 27, 629–640 (2008). https://doi.org/10.1109/TMI.2007.912817
4. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K. Van: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Trans. Med. Imaging. 34, 1993–2024 (2015). https://doi.org/10.1109/TMI.2014.2377694
5. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Trans. Med. Imaging. 35, 1240–1251 (2016). https://doi.org/10.1109/TMI.2016.2538465
6. Havaei, M., Dutil, F., Pal, C., Larochelle, H., Jodoin, P.-M.: A Convolutional Neural Network Approach to Brain Tumor Segmentation BT - Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Presented at the (2016)

7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation BT - Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Presented at the (2015)
8. Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J.: CE-Net: Context Encoder Network for 2D Medical Image Segmentation. IEEE Trans. Med. Imaging. 38, 2281–2292 (2019). https://doi.org/10.1109/TMI.2019.2903562
9. Dong, H., Yang, G., Liu, F., Mo, Y., Guo, Y.: Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks BT - Medical Image Understanding and Analysis. Presented at the (2017)
10. Kaewrak, K., Soraghan, J., Caterina, G.D., Grose, D.: Modified U-Net for Automatic Brain Tumor Regions Segmentation. In: 2019 27th European Signal Processing Conference (EUSIPCO). pp. 1–5 (2019)
11. Jiang, Z., Ding, C., Liu, M., Tao, D.: Two-Stage Cascaded U-Net: 1st Place Solution to BraTS Challenge 2019 Segmentation Task BT - Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Presented at the (2020)
12. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data. 4, 170117 (2017). https://doi.org/10.1038/sdata.2017.117
13. Bakas, S., Akbari, H., Sotiras, A.: Segmentation labels for the pre-operative scans of the TCGA-GBM collection. The Cancer Imaging Archive, (2017)
14. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. cancer imaging Arch. 286, (2017)
15. Lin, M., Chen, Q., Yan, S.: Network in network. arXiv Prepr. arXiv1312.4400. (2013)
16. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv Prepr. arXiv1412.6980. (2014)
17. Sim, K.S., Nia, M.E., Tso, C.P., Kho, T.K.: Chapter 34 - Brain Ventricle Detection Using Hausdorff Distance. In: Tran, Q.N. and Arabnia Bioinformatics, and Systems Biology, H.R.B.T.-E.T. in A. and I. for C.B. (eds.) Emerging Trends in Computer Science and Applied Computing. pp. 523–531. Morgan Kaufmann, Boston (2016)